**UNITED STATES DEPARTMENT OF COMMERCE**
**Bureau of the Census**
Washington, DC 20233-0001

December 7, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES R-38

MEMORANDUM FOR Maureen Lynch
Assistant Division Chief, Coverage Measurement Processing
Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: JEF for RC
Ryan Cromar and James Farber JEF
Sample Design Team

Subject: Accuracy and Coverage Evaluation: Final Large Block Cluster Subsampling Specifications

## I. INTRODUCTION

As the final stage of the A.C.E. sample design, large block cluster subsampling involves selecting a portion of a block cluster that has 80 or more A.C.E. housing units (HUs) to be in the A.C.E. interview sample. This will be accomplished by forming segments of adjacent HUs within the block cluster and selecting a subsample of segments. The objective of large block cluster subsampling is to meet the target A.C.E. interviewing sample sizes using the most up-to-date A.C.E. HU counts available at the time of subsampling. The large block cluster subsampling specification is similar to the specifications prepared for the 1998 Census 2000 Dress Rehearsal and documented in reference 1. The creation of the initial large block cluster subsampling parameter file, which is required to conduct large block cluster subsampling, is specified in reference 2.

These specifications also include instructions for assigning supplemental HUs to segments. Supplemental HUs are units found only in the census address list and not in the A.C.E. independent address list, and are not eligible for the A.C.E. interview sample. However, they must go through large block subsampling so that they are properly prepared for E-sample identification. A third task included in these specifications is the creation of interviewer workload assignments when the total number of HUs to interview in a block cluster exceeds 80.

Earlier stages of the A.C.E. sample design include the selection of A.C.E. block clusters for the listing sample (see reference 3), the A.C.E. reduction (see reference 4), and the

subsampling of small block clusters (see reference 5). After the listing sample selection, the independent listing is completed, the results keyed and verified, and the Independent List (IL) is created. Based on the results of this listing, the A.C.E. sample reduction and small block cluster subsampling are done. Subsequently, the HU matching and follow-up operations are done, and the preliminary Enhanced List (EL) is created and sent to large block cluster subsampling. The preliminary EL is both the input and output file for large block cluster subsampling. The output preliminary EL is updated with the results of subsampling, and is referred to as the subsampled preliminary EL. The Enhanced List is created by extracting only housing units designated for interview following large block cluster subsampling from the subsampled preliminary EL (see reference 6).

This memorandum is organized into the following sections:

- Assumptions
- Definitions
- Process Overview
- Input
- Process
- Output
- Verification
- References

Note that large block cluster subsampling is complete. This specification reflects the process as it was actually implemented. The original specification and the results of large block cluster subsampling are documented in references 7 and 8, respectively. This final specification includes changes to the original that were required due to issues that arose during production. In general, this final version is nearly identical to the original specification.

If there are any questions or comments, please contact James Farber (301-457-4282) or Deborah Fenstermaker (301-457-4195) of the Decennial Statistical Studies Division (DSSD).

## II. ASSUMPTIONS

The assumptions required for these specifications are given below.

A. Block clusters eligible for large block cluster subsampling include selected in the A.C.E. reduction and remaining in the sample following small block cluster subsampling. All other block clusters are not eligible for large block cluster subsampling.

B. For each block cluster in the listing sample, matching and HU follow-up operations have been completed and After Follow-up Match Codes have been assigned. Note that HUs in List/Enumerate and Relisted block clusters will not go through matching and follow-up but will have After Follow-up Match Codes assigned.

C. The A.C.E. HUs are the only HUs that are eligible for the A.C.E. interview sample. A.C.E. HUs are defined as HUs from the IL which have survived all HU follow-up procedures. These units have a match code of M, MU, UI, or CI, as defined in section III below. Note that even though future construction HUs were not included in the calculations of the initial sampling parameters, any future construction addresses that survived HU follow-up and matching are included as A.C.E. HUs.

D. Medium and small block clusters are eligible for large block cluster subsampling because the decision to subsample is based only on the number of A.C.E. housing units in the block cluster.

E. All A.C.E. HUs on the IL are keyed and valid.

F. All decimal numbers are rounded to six digits at the time of creation using the standard rounding procedure, except when noted otherwise. Decimal numbers with a seventh decimal place of five or more are rounded up in the sixth decimal place. Those with four or less in the seventh decimal place are rounded down in the sixth decimal place.

G. The initial large block cluster subsampling parameter file has been created and verified.

H. Large block cluster subsampling is done on a flow basis over a span of several days. Therefore, daily large block cluster subsampling parameter files will be created and will include daily statistics to track the day-to-day results of large block cluster subsampling. These files will also provide the needed inputs for processing on successive days. Large block cluster subsampling may run

concurrently with operations such as HU follow-up and Target Extended Search (TES) sampling.

I. There will be no large block cluster subsampling in American Indian Reservation (AIR) clusters, using the AIR definitions known at the time of block clustering.

## III. DEFINITIONS

### A. After Follow-up Match Code

A code assigned to HUs after HU follow-up. For the purposes of these specifications, the only match codes that need to be defined are those that occur on the preliminary EL. As documented in references 6 and 9, these match codes are:

- M = The A.C.E. and census addresses match.
- MU = The A.C.E. and census addresses match and there is not enough information on the follow-up form to confirm this match as a HU with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.
- UI = Not enough information on the follow-up form to assign a code to the nonmatched A.C.E. HU with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.
- UE = Not enough information on the follow-up form to assign a code to the census nonmatched HU with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.
- CI = The A.C.E. housing unit existed as a HU at the time of the follow-up interview and is correctly geocoded in the block cluster. The HU is not found in the census.
- CE = The census housing unit existed as a HU at the time of the follow-up interview and is correctly geocoded in the block cluster. The HU is not found in the A.C.E..

### B. American Indian Reservation Block Cluster

A block cluster at least partially on an AIR with three or more HUs based on information available at the time of block clustering. See Sampling Strata.

C. Block Cluster

A geographically contiguous group of Census 2000 collection blocks (see reference 10).

D. A.C.E. Housing Unit

A housing unit on the preliminary EL that is keyed and valid and has one of the following After Follow-up Match Codes: M, MU, UI, or CI.

E. Housing Unit Follow-up

Reconciliation in the field of HUs not matched in the matching process.

F. A.C.E. Independent List

List of all HUs in A.C.E. listing sample clusters. The IL is created independently of the Decennial Master Address File (DMAF), the census address list.

G. Keyed and Valid HUs

A.C.E. HUs with one of the following Unit Status codes:

1 = Occupied or vacant and intended for occupancy
2 = Under construction
3 = Future construction
4 = Unfit for habitation
5 = Boarded up
6 = Storage of household goods
7 = Vacant mobile home site
8 = Other

All of these units are included because it is possible that the unit status may change between listing and interviewing. Group quarters are not listed in A.C.E.

H. Large Block Cluster

See Sampling Strata.

I. Listing Sample

The initial sampling stage of the A.C.E. survey in which block clusters are selected for independent listing (see reference 3).

J. Matching

Computer and clerical process of comparing the IL and the DMAF to determine which addresses are common to both lists and which are on only one list.

K. Medium Block Cluster

See Sampling Strata.

L. Preliminary Enhanced List

The input and output of the large block cluster subsampling process. The output is referred to as the subsampled preliminary enhanced list.

M. A.C.E. Reduction

The process of reducing the A.C.E. listing sample from the Integrated Coverage Measurement (ICM) sample to the A.C.E. interview sample. In the A.C.E. reduction, the listing sample block clusters are subsampled, and the selected clusters continue to small block cluster subsampling (see Reference 4).

N. A.C.E. Reduction Strata

A partition (mutually exclusive and exhaustive set) of all block clusters in a state into groups according to certain characteristics. See Attachment A for a list of the A.C.E. Reduction Strata and see reference 4 for more information on how A.C.E. reduction strata were defined.

O. Sampling Strata

A partition of all block clusters within a state into groups according to the number of HUs estimated in each cluster at the time of block clustering (see reference 11). Sampling strata were assigned to block clusters prior to listing sample selection. The sampling strata are:

| | | |
|---|---|---|
| 1 | = | small block clusters with 0 - 2 estimated HUs |
| 2 | = | medium non-AIR block clusters with 3 - 79 estimated HUs |
| 3 | = | large non-AIR block clusters with $\geq$ 80 estimated HUs |
| 4 | = | medium and large AIR block clusters with $\geq$ 3 estimated HUs |

Note that medium and even small block clusters are eligible for large block cluster subsampling because the decision to subsample is based only on the number of A.C.E. HUs in the block cluster.

P. Small Block Cluster

See Sampling Strata.

Q. State

In this specification, a state is one of the 50 United States plus the District of Columbia and Puerto Rico.

R. Supplemental Housing Unit

A housing unit found in the version of the DMAF used for housing unit matching but not found in the IL. Supplemental HUs are not A.C.E. HUs and are not eligible for the A.C.E. interview sample. However, they are assigned to segments during large block cluster subsampling to facilitate E-sample identification (see reference 12). The After Follow-up Match Codes for the supplemental HUs are CE and UE.

## IV. PROCESS OVERVIEW

Large block cluster subsampling is used to achieve the target A.C.E. interview sample size in each A.C.E. reduction stratum within each state. These target sample sizes are designed to increase the weights of large block clusters and minimize weight variation between medium and large block clusters in the same reduction stratum within a state. This overview will detail the steps of the large block cluster subsampling process. The steps listed here correspond to the steps in section VI below, which contains the programming instructions and is significantly less descriptive than this overview.

A. Create Block Cluster HU Counts and Subsampling Status Codes

The preliminary EL includes not only A.C.E. HUs but also supplemental HUs, and therefore a step is required to determine the various HU counts. This step also determines if large block cluster subsampling is required in a cluster. For each cluster, the following HU counts are computed:

- total number of preliminary EL HUs
- total number of A.C.E. HUs
- total number of supplemental HUs

Using these counts and the large block cluster subsampling take-every (TELB) from the parameter file, the following table determines if large block cluster

subsampling is required. Large block cluster subsampling occurs only in certain block clusters, as shown in Table 1:

Table 1. Where Large Block Cluster Subsampling is Required

| # A.C.E. HUs | TELB | In AIR? | Lg. BC Subsampling Required? |
|---|---|---|---|
| Less than 80 | 1.000000 | Yes | No |
| Less than 80 | 1.000000 | No | No |
| 80 or more | 1.000000 | Yes | No |
| 80 or more | 1.000000 | No | No |
| 80 or more | >1.000000 | No | Yes |

Table 1 shows that only block clusters that have 80 or more A.C.E. HUs, have a TELB greater than one, and are not in an AIR go through large block cluster subsampling. There is no subsampling on an AIR regardless of the size of the block cluster. Nor is there subsampling if the number of supplemental HUs exceeds a certain number. Note that sampling stratum also does not affect the need for subsampling. A block cluster that was originally designated as small or medium may end up with 80 or more A.C.E. HUs after independent listing, and therefore would go through large block cluster subsampling.

Clusters that do not need large block cluster subsampling skip to step E. Clusters that require subsampling continue to step B.

B. Create Segments

Only non-AIR clusters with 80 or more A.C.E. HUs and a TELB greater than one need to be segmented for large block cluster subsampling. The number of segments to form in clusters in each A.C.E. reduction stratum within each state was determined during the creation of the initial large block cluster subsampling parameter file. However, the size of the segments, the number of A.C.E. HUs in each segment, will vary among the clusters in the same reduction stratum within a state. Determining the segment size for a cluster is the first step in the creation of segments. The segment size calculation is simply the number of A.C.E. HUs divided by the prespecified number of segments. Because this calculation will usually not yield an integer sample size, an algorithm is used to distribute the remainder among segments.

After the segment sizes for a cluster are determined, the A.C.E. HUs are assigned to segments. The A.C.E. HUs are assigned to the first segment until that

segment's size is reached. Then they are assigned to the second segment until its target size is reached, and so on until all segments attain their designated sizes.

The final step of segment creation is the assignment of supplemental HUs to segments. Supplementals are not eligible for an A.C.E. interview, but they still require a segment identifier to facilitate E-sample identification. Generally, supplemental HUs are assigned to the same segment as the nearest preceding A.C.E. HU. However, it is possible for a supplemental HU to occur before all A.C.E. HUs in a cluster, in which case the supplemental HU goes into the first segment in the cluster.

C. Create Segment Level Variables and Codes

For each segment, total HU counts similar to those in step A above need to be computed. The HU counts required are the total number of HUs, the total number of A.C.E. HUs, and the total number of supplemental HUs. These counts are used to compute interviewing sample sizes after subsampling.

D. Select a Sample of Segments for Each A.C.E. Reduction Stratum within State

This is the actual subsampling step. A systematic sample of segments is selected within each A.C.E. reduction stratum and state for inclusion in the A.C.E. interview sample. A complication of this subsampling operation is that clusters arrive into this step on a flow basis. Ideally, the subsampling of segments would occur as a single operation after all matching and follow-up operations have been done and all clusters have been placed on the preliminary EL. However, it is essential that the interview sample get to the field as quickly as possible, and thus large block cluster subsampling will be performed on a daily basis until all eligible clusters are subsampled. Despite the daily processing, the subsampling is designed as if it was a single operation, where the universe is all clusters available on a given day instead of all clusters in the A.C.E. sample. The only loss is the ability to sort all clusters together. Instead, a geographic sort of clusters available on each day will be done to minimize geographic bias in the interview sample.

A sample of segments is selected from all clusters available for subsampling on a given day within each A.C.E. reduction stratum and state. At the end of each day, the point where the sampling ends is carried over as the starting point for the next day, so only one random start is required for each A.C.E. reduction stratum within each state. Subsampling data from each day is saved on the large block cluster subsampling parameter file for verification purposes. The sample of segments is selected using the standard systematic sampling technique with a random start.

E. Identify Housing Units for the A.C.E. Interview Sample

After segments have been selected for the A.C.E. interview sample, the appropriate HUs in those segments need to be identified for interviewing. Only A.C.E. HUs in selected segments are in the A.C.E. interview sample. Supplemental HUs in selected segments are not designated for interview. An interview flag is set on the preliminary EL for A.C.E. HUs designated for interviewing. A later operation extracts the HUs to be interviewed from the subsampled preliminary EL, and the resulting file is the Enhanced List.

F. Create Interviewer Workload Assignments

This step occurs only for clusters with 80 or more interview HUs. In these clusters, interviewer assignments of 40 to 50 HUs will be created to facilitate interviewing efficiency. Assignment workloads will be balanced such that the maximum difference between any two assignments in a cluster will be one HU. The maximum number of assignment workload areas in a single cluster is 26.

G. Update or Create Files

Files such as the Sample Design File and the Large Block Subsampling Segment File will be updated throughout the process to facilitate subsampling verification. At the end of each day, the Large Block Cluster Subsampling Parameter File will also be updated to provide information needed to begin processing on the next day. The final versions of these files will be created at the end of large block cluster subsampling. Note that one of the results of large block cluster subsampling will be the A.C.E. weight for each interviewed HU. This weight is computed as the product of the take-everys from all previous sampling operations.

This step also updates files for clusters that have zero HUs and thus were not eligible for large block cluster subsampling.

V. INPUT

The input sources for the large block cluster subsampling process are the following:

A. Initial Large Block Cluster Subsampling Parameter File

Description: This file contains sampling parameters needed for selecting the systematic sample on a flow basis. The final version will be created when production is complete. The creation of this file is documented in reference 2.
Level: A.C.E. Reduction Stratum
Scope: One record per A.C.E. reduction stratum within each state
Layout: See Attachment A

B. Cluster Status File

Description: This file has one record for each block cluster selected for the A.C.E. listing sample. It is updated with information from other processing stages. For large block cluster subsampling, this file is used to determine the subsampling parameters and to determine when a cluster is available for subsampling.
Level: Block Cluster
Scope: One record for each cluster in the A.C.E. listing sample

C. Daily Large Block Cluster Subsampling Parameter File

Description: This file contains information for selecting the systematic sample on a flow basis. The file will be produced after the sampling parameters are calculated and will be updated daily during large block cluster subsampling to record the starting point for the next day's systematic sampling. The final version will be created when production is complete.
Level: A.C.E. Reduction Stratum
Scope: One record per A.C.E. reduction stratum within each state
Layout: See Attachment A

D. Preliminary Enhanced List

Description: The file is created from the matching of the IL and the DMAF and the associated housing unit follow-up of non-matches. The types of HUs on this file are IL-only, IL and DMAF, and supplementals (DMAF-only).
Level: Housing Unit
Scope: One record for each HU in clusters selected for A.C.E. following small block cluster subsampling
Layout: See Attachment B

E. A.C.E. Sample Design file (Version 3)

Description: This file reflects the previous A.C.E. sampling operations: listing sample selection, A.C.E. reduction, and small block cluster subsampling.
Level: Block Cluster
Scope: One record for each cluster in the A.C.E. listing sample
File Layout: See Attachment C

## VI. PROCESS

The following are the steps of large block cluster subsampling. See section IV above for a detailed overview of these steps. These steps are completed on a flow basis for each cluster that remains in the A.C.E. sample following small block cluster subsampling. For those clusters that have fewer than 80 A.C.E. HUs, the process is simple because no segmenting or subsampling is required. The process is more complex for clusters that have 80 or more A.C.E. HUs. Attachment D gives an example of the process, and Attachment E provides a flowchart.

A. Create Block Cluster Housing Unit Counts and Subsampling Status Codes

Create 10 variables for each block cluster regardless of its size to track processing.

1. Determine HU counts from the preliminary EL in each cluster as follows:

   a. Count the total number of HUs, NHUEL.

   b. Count the total number of A.C.E. HUs, NHUELA.

   c. Count the total number of supplemental HUs, NHUELN.

d. Calculate Z to check the counts above.

$$Z = NHUEL - (NHUELA + NHUELN)$$

If the counts are correct, Z will be equal to zero. Resolve the cases where Z is not equal to zero.

2. Determine TELB and RSLB as follows:

- If NHUELA < 80, then set TELB and RSLB equal to 1.000000.
- If NHUELA ≥ 80, then set TELB and RSLB equal to the TELB and RS values obtained from the large block cluster subsampling parameter file for the corresponding state and reduction stratum.

3. Use the HU counts calculated above, TELB, and the AICIND variable from the Sample Design File Version 3 to assign the subsampling status codes below according to Table 2:

- ELLBSUB is a block cluster flag to denote if subsampling is required, and is assigned to the entire cluster
- NSEGSAM is the number of segments in sample in the cluster, and is also assigned to the entire cluster
- SEGSUB is the flag that indicates if a segment is in sample or not, and is assigned to individual segments within a cluster
- SEGID is a two-character code identifying each segment and the HUs in each segment, and is assigned to segments and HUs

Note that subsampling status codes are unknown at this step for non-AIR clusters with 80 or more A.C.E. HUs and a TELB greater than one.

Table 2 also indicates to which step a cluster should proceed for each possible subsampling status. Clusters that do not require subsampling proceed to the interview sample identification. Clusters that do require subsampling go through the segmenting and subsampling processes before the interview sample can be identified.

Table 2. Summary of Subsampling Status Codes

| Value of NHUELA | TELB | AICIND | ELLBSUB | NSEGSAM | SEGSUB | SEGID | Go to Step |
|---|---|---|---|---|---|---|---|
| less than 80 | 1.000000 | 1 | 1 | 1 | 1 | AA | E |
| less than 80 | 1.000000 | 0 or 2 | 1 | 1 | 1 | AA | E |
| 80 or more | 1.000000 | 1 | 1 | 1 | 1 | AA | E |
| 80 or more | 1.000000 | 0 or 2 | 1 | 1 | 1 | AA | E |
| 80 or more | >1.000000 | 0 or 2 | 2 | Unknown | Unknown | Unknown | B |

B. Create Segments

Only non-AIR clusters with 80 or more A.C.E. HUs and a TELB greater than one need segments. For each cluster that requires segmenting, do the following:

1. Calculate the number of A.C.E. HUs in each segment as follows:

   a. Determine the number of segments to form from the variable NSEG on the initial large block cluster subsampling parameter file. If NSEG = 1, assign SEGID = AA to all HUs in the cluster and proceed to step E below and treat the block cluster as if it had fewer than 80 A.C.E. HUs or was on an AIR.

   b. Compute the average number of A.C.E. HUs per segment as
   $$\frac{NHUELA}{NSEG}.$$

   c. Truncate this number to an integer, and denote the truncated value AVGSEG. Compute the number of remaining A.C.E. HUs, R, as R = NHUELA - (NSEG × AVGSEG)

   d. Assign SEGSIZE for each segment as follows:

      - For the first R segments, SEGSIZE = AVGSEG + 1
      - For the remaining NSEG - R segments, SEGSIZE = AVGSEG

2. Within each cluster, assign A.C.E. HUs to segments as follows:

   a. Sort all preliminary EL HUs (A.C.E. and supplemental HUs) by map spot number (MSN) and within-map spot number identification (WMSN).

b. Assign SEGID and A.C.E. HUs to the NSEG segments. The first segment in the cluster receives SEGID AA. Assign A.C.E. HUs to this segment until it contains its predetermined number of A.C.E. HUs as indicated by its SEGSIZE. The second segment is then given a SEGID of BA, and A.C.E. HUs are assigned to segment BA until its value of SEGSIZE is reached. Continue with segment CA and so forth in the same manner until all NSEG segments contain as many A.C.E. HUs as their values of SEGSIZE dictate. If there are more than 26 segments in a cluster, continue SEGID as AB, BB, CB, and so forth. Assign SEGID values for each A.C.E. HU in the cluster the same value as SEGID for their segment.

3. Assign supplemental HUs to the same segment as the nearest preceding A.C.E. HU. If supplemental HUs occur before the first A.C.E. HU in a cluster, set SEGID = AA for these supplemental HUs.

C. Create Segment Level Variables and Codes

Determine HU counts in each segment as follows:

1. Count the total number of HUs in the segment, NHUELS.

2. Count the number of A.C.E. HUs in the segment, NHUELAS.

3. Count the number of supplemental HUs in the segment, NHUELNS.

4. Calculate Z to check the counts above:

$$Z = NHUELS - (NHUELAS + NHUELNS)$$

If the counts are correct, Z will be equal to zero. Resolve the cases where Z is not equal to zero.

D. Select a Sample of Segments within A.C.E. Reduction Stratum and State

Do the following sampling procedure separately for each A.C.E. reduction stratum within each state. Sort the clusters available on a given day by A.C.E. cluster number within reduction stratum within state before subsampling.

1. Create the variable DAY to record the day on which a cluster arrives after HU matching and follow-up. When the first cluster arrives, set DAY = 1. Increment DAY by one on each following day until all clusters have

arrived. On a particular day, assign the same DAY value to all clusters processed on that day.

2. Create the cluster-level variable CON to identify the order in which the clusters are arranged within a particular A.C.E. reduction stratum and state prior to subsampling:

   a. Obtain the cumulative cluster count, CCC, from the large block cluster subsampling parameter file from the previous day.

   b. Set CON = CCC + 1 for the first cluster processed on a given day. Increment CON by one for each remaining cluster to be processed on that day.

   c. When the last cluster on a given day has been processed, set CCC = CON for the last block cluster, and save CCC to the current day's parameter file.

3. Create the segment level variable DSON to identify the order in which the segments are arranged within clusters in a particular reduction stratum and state on a single day as follows:

   a. Set DSON = 1 for segment AA in the first cluster in the first A.C.E. reduction stratum within state to be subsampled that day.

   b. Increment DSON by one for each remaining segment in all clusters in that same A.C.E. reduction stratum within a state to be subsampled that day.

   c. Let N = the maximum value for DSON for the A.C.E. reduction stratum and state on that day.

4. Select a systematic subsample of segments for each state and A.C.E. reduction stratum:

   Generate a sequence of numbers $L_1, ..., L_n$ as follows:

   a. Obtain the current daily start value, DS, and the take-every for the current A.C.E. reduction stratum within each state, TELB, from the current day's parameter file.

   b. Let $L_1 = DS$.

c. Calculate $L_j = L_{j-1} + TELB$, for $j = 2, ..., n$, where n is the largest integer such that $[DS + (n - 1) \times TELB] \leq N$.

d. Round each $L_j$ up to the nearest integer (an integer rounds to itself).

e. For each segment in the reduction stratum with a DSON equal to the rounded values of $L_j$, $j = 1, ..., n$, assign SEGSUB = 1. These segments are in the A.C.E. interview sample.

f. For each segment in the sampling stratum with a DSON not equal to the rounded values of $L_j$, $j = 1, ..., n$, assign SEGSUB = 0. These segments are not in the sample.

g. Calculate the daily end value, $DE = DS + (TELB \times n) - N$.

h. Save DE as DS on the next day's sampling parameter file.

i. To monitor the day-to-day sampling progress, keep track of the daily start value. For each day of sampling, update DS on the large block cluster subsampling parameter file. Name these variables according to the day of sampling (e.g., DS1 is the daily start for day 1, DS2 is the daily start for day 2, ..., DS20 is the daily start for the final day[1]). If no sampling occurs in a reduction stratum for a particular day, the corresponding variable will be blank.

For example:

On day one, if N = 40, TELB = 4.500000, and RS = DS = 1.553000, then n = 9. Set $L_1 = 1.553000$. The generated $L_j$s would be the sequence: 1.553000, 6.053000, 10.553000, 15.053000, 19.553000, 24.053000, 28.553000, 33.053000, and 37.553000. Therefore, the segments with DSON values of 2, 7, 11, 16, 20, 25, 29, 34, and 38 would be selected for the sample. The daily end is $DE = 1.553000 + 4.500000 \times 9 - 40 = 2.053000$.

On day two, if N = 15, TELB = 4.500000, and DS = 2.053000, then n = 3. Set $L_1 = 2.053000$. The generated $L_j$s would be the sequence: 2.053000, 6.553000, and 11.053000. Therefore, the segments with DSON values of 3, 7, and 12 would be selected for the sample. The daily end is $DE = 2.053000 + 4.500000 \times 3 - 15 = 0.553000$. This continues until all block clusters have been processed.

[1]The number of days for sampling may be over or under 20. If this is the case, make appropriate changes.

5. Check the number of sampled segments daily by calculating c:

$$c = \left| \frac{N}{TELB} - n \right|$$

If the sampling is implemented correctly, c will be less than one. For values of c that are not less than one and have not been resolved, contact the Sample Design Team for review of the sampling operations.

6. For each block cluster, count the number of segments selected for the A.C.E. interview sample, NSEGSAM.

E. Identify Housing Units for the A.C.E. Interview Sample

All A.C.E. HUs in selected segments will be sent to interview. All A.C.E. HUs in non-selected segments and supplemental HUs in all segments will not be sent to interview.

1. Note that clusters that did not undergo subsampling rejoin the process at this point. These clusters and their A.C.E. HUs already have SEGSUB, NSEGSAM, and SEGID values assigned. Assign the following fields for these clusters and their HUs:

   - DAY = 1 for the cluster
   - SEGSIZE = NHUELA for the one segment in the cluster
   - NHUELS = NHUEL for the segment
   - NHUELAS = NHUELA for the segment
   - NHUELNS = NHUELN for the segment
   - CON = Blank for the A.C.E. HUs in the cluster
   - DSON = Blank for the A.C.E. HUs in the cluster
   - Assign SEGID = AA for all supplemental HUs in the cluster

2. For all clusters, regardless of whether or not they were subsampled, create an A.C.E. interview flag for each A.C.E. and supplemental HU, INTERVW, and assign as follows:

   - If SEGSUB = 1 then
     INTERVW = 1 for all A.C.E. HUs
     INTERVW = 9 for all supplemental HUs
   - If SEGSUB = 0, then
     INTERVW = 0 for all A.C.E. HUs
     INTERVW = 8 for all supplemental HUs

3. Compute interview HU counts for use in forming workload assignments:

   a. Count the number of total HUs for interview in each cluster, NINT

   b. Count the number of total interview HUs in each segment, NINTS

F. Create Interviewer Workload Assignments

Create interviewer workload assignments in clusters of 80 or more interview HUs:

1. Calculate the number of assignments needed for each cluster, NA, and the size of the assignments, ASIZE, as follows:

   - If NINT < 80, then
     NA = 1
     ASIZE = NINT

   - If NINT ≥ 80, then

     a. Compute $NA = \frac{NINT}{40}$. If NA is not an integer, round it down to the next integer. If NA > 26, set NA = 26.

     b. Calculate the average number of HUs per assignment, AVGHUA, as follows:

     $$\frac{NINT}{NA}.$$

     Truncate this number to an integer. AVGHUA is the truncated value.

     c. Calculate the number of remaining HUs for interview, RINT:

     $$RINT = NINT - (NA \times AVGHUA)$$

     d. Calculate ASIZE for each assignment as follows:

     - For the first RINT assignments,
       ASIZE = AVGHUA + 1

     - For the remaining NA - RINT assignments,
       ASIZE = AVGHUA

2. Assign the interviewer assignment identifier to interview HUs and supplemental HUs in each sampled segment as follows:

   Create an assignment identifier, ASSIGNID, to distinguish among assignments within a cluster.

   - If NINT < 80, then ASSIGNID = AA for all interview HUs and supplemental HUs in the cluster
   - If NINT ≥ 80, then

     a. Sort the interview HUs by MSN and WMSN in the block cluster.

     b. Assign ASSIGNID and interview HUs to the NA assignments. The first assignment in the block cluster receives ASSIGNID AA. Assign interview HUs to this assignment until it contains its predetermined number of interview HUs as indicated by its ASIZE. The second assignment is then given an ASSIGNID of AB, and interview HUs are assigned to assignment AB until its value of ASIZE is reached. Continue with assignment AC and so forth in the same manner until all NA assignments contain as many interview HUs as their values of ASIZE dictate. The maximum value of ASSIGNID is AZ since no more than 26 assignments can be created in a block cluster.

     c. Assign supplemental HUs in sampled segments the same ASSIGNID as the nearest preceding interview HU. If supplemental HUs occur before the first A.C.E. HU in a cluster, set ASSIGNID = AA for these supplemental HUs.

     d. Use ASSIGNID for interview HUs (INTERVW = 1) to distinguish workload assignments on the preliminary EL.

G. Update or Create Files

1. Daily and Final Large Block Cluster Subsampling Parameter Files

   At the end of each day, update the daily large block cluster subsampling parameter file for that day with the following information:

   Daily Start, DS
   Cumulative Cluster Count, CCC
   Daily Start for the Current Day *i*, DS*i*

On the final day of processing, create the final large block cluster subsampling parameter file by copying the final day's parameter file.

2. Preliminary Enhanced List

Append the following HU variables to the preliminary EL during the large block cluster subsampling process:

Segment Identifier, SEGID
Assignment Identifier, ASSIGNID
Interview Flag, INTERVW

3. Daily and Final Large Block Cluster Subsampling Segment Files

Create a file each day that includes the segments created in clusters that have more than one segment. Include the following in these files:

| Variable Description | Name | Char |
|---|---|---|
| State | STATE | 1-2 |
| County | CTY | 4-6 |
| A.C.E. Reduction Stratum | ARST | 8-9 |
| Sampling Stratum | SS | 11 |
| A.C.E. Block cluster number and Check Digit | CLUST | 13-18 |
| Day of Arrival | DAY | 20-21 |
| Segment Identifier | SEGID | 23-24 |
| Daily Segment Order Number | DSON | 26-28 |
| Number total HUs in segment | NHUELS | 30-34 |
| Number A.C.E. HUs in segment | NHUELAS | 36-40 |
| Number supplemental HUs in segment | NHUELNS | 42-46 |
| Segment subsampling code | SEGSUB | 48 |
| Number of HUs for interview in segment | NINTS | 50-54 |

On the final day of processing, create a final large block cluster subsampling segment file by concatenating the segment files for all of the days into a single file.

4. Daily Sample Design Files

Update version three of the Sample Design File with the results of large block cluster subsampling for the first day and then on each successive day update the preceding day's Daily Sample Design File by appending the following cluster-level information. Create the variables RELIST and WEIGHTP for the Sample Design File:

- RELIST: Relisted block cluster flag
  0 = Block cluster not relisted
  1 = Block cluster relisted

- WEIGHTP: Block cluster A.C.E. weight
  WEIGHTP = TELB x TE1 x TE2 x TEAR x FTESB,
  where TE1, TE2, TEAR, and FTESB are obtained from the Sample Design File Version 3.

- FSS: Final Sampling Stratum
  Concatenate the following variables to create FSS:
  - State, ST
  - Small Block Cluster Subsampling Stratum, SBCSS
  - A.C.E. Reduction Stratum, ARST

| Variable Description | Name | Char |
|---|---|---|
| Relisted block cluster flag | RELIST | 371 |
| Number of total HUs in block cluster | NHUEL | 373-377 |
| Number of A.C.E. HUs in block cluster | NHUELA | 379-383 |
| Number of supplemental HUs in block cluster | NHUELN | 385-389 |
| Large block cluster subsampling code | ELLBSUB | 391 |
| Random Start for large block cluster subsampling | RSLB | 393-403 |
| Take-every for large block cluster subsampling | TELB | 405-415 |
| Number of segments in block cluster | NSEG | 417-418 |
| Number of segments selected in block cluster | NSEGSAM | 420-421 |
| Day of Arrival for block cluster | DAY | 423-424 |
| Final Cluster Order Number | CON | 431-434 |
| Number of total HUs for interview in block cluster | NINT | 436-440 |
| Unbiased Weight for P-sample HUs | WEIGHTP | 442-453 |
| Number of Assignments in block cluster | NA | 455-456 |
| Final Sampling Stratum | FSS | 458-464 |

5. Sample Design File, Version 4

After the last day of large block cluster subsampling, create version 4 of the Sample Design File as follows:

a. For clusters that went through large block cluster subsampling, including those that did not require segmenting or subsampling, the Sample Design File will have been updated during the process so no further updates are required.

b. For block clusters in the A.C.E. sample (CSI = 1 on the Sample Design File) that have zero total HUs, assign values to variables as follows and include these fields in version 4 of the Sample Design File using the layout in section VI.G.4 above.

   - RELIST, WEIGHTP, and FSS are defined as above

- Set the following fields to one: ELLBSUB, NSEG, NSEGSAM, DAY, NA, RSLB, TELB (RSLB and TELB are decimal numbers with six digits after the decimal)
- Set the following fields to zero: NINT, NHUEL, NHUELA, NHUELN
- Set CON to Blank

c. For clusters not in the A.C.E. sample (CSI = 0 on the Sample Design File), blank all fields listed in section VI.G.4 above.

d. Because the TES sample selection was done concurrently with large block cluster subsampling, the variables created by the TES process will also be in Version 4 of the Sample Design File.

## VII. OUTPUT

The outputs requested by the Sample Design Team are the following:

A. Daily and Final Large Block Cluster Subsampling Parameter Files

See section V for the description of these files and Attachment A for the layout of these files.

B. Subsampled Preliminary Enhanced List

Description: The subsampled preliminary EL is the input preliminary EL updated with the results of large block cluster subsampling. All HUs in A.C.E. sample block clusters are on the subsampled preliminary EL. The EL will be created by extracting the HUs designated for interview from the subsampled preliminary EL. The subsampled preliminary EL also provides the input for E-sample identification.

Level: Housing Unit

Scope: One record for each HU in clusters selected for A.C.E. following small block cluster subsampling.

Layout: See Attachment B. Additions are:

Segment Identifier
Assignment Identifier
Interview Flag

C. Sample Design File (Version 4 - Daily and Final)

Description: This file reflects the sampling through large block cluster subsampling and TES. This file will be produced on a daily basis during large block cluster subsampling. The final version will be created when production is complete.
Level: Block Cluster
Scope: All block clusters selected in the initial A.C.E. sampling
File Layout: See Attachment C

D. Large Block Cluster Subsampling Segment File (Daily and Final)

Description: This file contains segment level information for all segments in all A.C.E. block clusters. This file will be produced on a daily basis during large block cluster subsampling. The final version will be created when production is complete.
Level: Block cluster segment
Scope: All block clusters that have more that one segment created during large block cluster subsampling.
File Layout: See section VI.G.3 above

## VIII. VERIFICATION

The following information should be provided for verification:

A. Large Block Cluster Subsampling Parameter File

Provide the sampling parameter file. The initial version of this file contains variables calculated from HU totals on the IL. Therefore, the initial version can be created and verified prior to forming and subsampling segments. Provide subsequent daily versions of this file during the large block cluster segment subsampling. See part A in section VII for the description of this file and Attachment A for the layout of this file.

B. Preliminary Enhanced List File

Make available the preliminary EL. Using this file, the Sample Design Team will verify HU totals, the assignment of subsampling status codes, segment creation, and workload assignments. Make this file available daily.

C. Updated Sample Design File and Large Block Cluster Segment File

Provide an updated version of the Sample Design File and the large block cluster segment file daily, and provide the final versions of these files at the end of the process. Using these files in conjunction with the sampling parameter file, DSSD will verify the implementation of the daily subsampling process.

D. Cluster Status File

Make available the Cluster Status File daily. Using this file, the Sample Design Team will verify that clusters with zero total HUs have large block cluster subsampling values assigned properly in the Sample Design File.

## IX. REFERENCES


1 DSSD Census 2000 Dress Rehearsal Memorandum Series A-9, "Census 2000 Dress Rehearsal ICM Sampling: Large Block Subsampling Specification," April 15, 1998.

2 DSSD Census 2000 Procedures and Operations Memorandum Series R-26, "Accuracy and Coverage Evaluation: Large Block Cluster Subsampling Parameter File Specification," March 8, 2000.

3 DSSD Census 2000 Procedures and Operations Memorandum Series R-3, "Accuracy and Coverage Evaluation (A.C.E.) Survey: Block Cluster Sample Selection Specification," March 29, 1999.

4 DSSD Census 2000 Procedures and Operations Memorandum Series R-29, "Accuracy and Coverage Evaluation Survey: Reduction Specification," March 21, 2000.

5 DSSD Census 2000 Procedures and Operations Memorandum Series R-24, "Accuracy and Coverage Evaluation Survey: Small Block Cluster Subsampling," February 1, 2000.

6 DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-HU-08, "Creation of the Census 2000 Accuracy and Coverage Evaluation (A.C.E.) Enhanced List for Person Phase Interviewing," December 5, 2000.

7 DSSD Census 2000 Procedures and Operations Memorandum Series R-27, Accuracy and Coverage Evaluation: Large Block Cluster Subsampling Specifications," March 8, 2000.

8 DSSD Census 2000 Procedures and Operations Memorandum Series R-32, Accuracy and Coverage Evaluation Survey: Large Block Cluster Subsampling Approval and Summary of Results," May 30, 2000.

9 DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-01, "Accuracy and Coverage Evaluation: The Design Document," January 11, 2000.

10 DSSD Census 2000 Procedures and Operations Memorandum Series R-8, "Census 2000 Specifications for Block Cluster Formation-Reissue," May 3, 1999.

11 DSSD Census 2000 Procedures and Operations Memorandum Series R-4, "Accuracy and Coverage Evaluation (A.C.E.) Survey: Sample Summary File and Sample Design File Documentation," March 30, 1999.

12 DSSD Census 2000 Procedures and Operations Memorandum Series R-31, "Accuracy and Coverage Evaluation Survey: Specification for E-Sample Identification," May 16, 2000.

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
A.C.E. Implementation Team/Statistical Design Team Leaders List
Sample Design Team

Layout of Sampling Parameter Files

The Initial Daily and Final Large Block Cluster Subsampling Parameter Files have the following file layout:

| Variable Description | Name | Pos |
|---|---|---|
| State | ST | 1-2 |
| A.C.E. reduction stratum | ARST | 4-5 |
| Target housing unit sample size | T | 7-14 |
| Number of housing units in block clusters with 80 or more housing units on the independent list | NILHUL | 16-21 |
| Number of housing units in block clusters with 0-79 housing units on the independent list (except smalls with 0-9) | NILHUM | 23-28 |
| Number of housing units in all block clusters on the independent list | NILHUT | 30-35 |
| Number of housing units in small block clusters with 0-9 housing units on the independent list | NILHUS | 37-42 |
| Take-every for the segment subsampling | TELB | 44-54 |
| Number of segments in a block cluster | NSEG | 56-57 |
| Flag for formula used for calculating NSEG | FORMULA | 59 |
| Random Number between 0 and 1 | RN | 61-72 |
| Random Start for the segment subsampling | RS | 74-84 |
| Current Daily Start | DS | 86-96 |
| Cumulative Cluster Count | CCC | 98-100 |
| Daily Start for Day 1 | DS1 | 102-112 |
| Daily Start for Day 2 | DS2 | 114-124 |
| . | . | . |
| . | . | . |
| . | . | . |
| Daily Start for Day 20[2] | DS20 | |

[2]The number of days for sampling may be over or under 20. If this is the case, appropriate modifications will be made.

# Layout of the Preliminary Enhanced List

Layout Name : ENHANCED00.LAY Page : 1
Description : 2000 ENHANCED LIST LAYOUT
Total Length : 360
Date Created : 11-22-1999

| # | Field | Field description | length | Positions Beg - End | |
|---|---|---|---|---|---|
| 1. | CNTRLNM | CONTROL NUMBER<br>1: 4 LCO<br>5:10 CLUSTER<br>11:12 SEGMENT<br>13:17 MAP SPOT NUMBER<br>18:21 WITHIN MSN ID<br>22:24 ZERO FILL | 24 | 1 - 24 | CHAR |
| 2. | LCO | LOCAL CENSUS OFFICE | 4 | 25 - 28 | CHAR |
| | | ********** Index 1 CLUST thru WMSN ********** | | | |
| 3. | CLUST | CLUSTER NUMBER | 6 | 29 - 34 | CHAR |
| 4. | MSN | ENHANCED IL MAP SPOT NUMBER | 5 | 35 - 39 | CHAR |
| 5. | WMSN | WITHIN MAP SPOT NUMBER ID | 4 | 40 - 43 | CHAR |
| | | ********** Index 2 CID ********** | | | |
| 6. | CID | MAF ID | 12 | 44 - 55 | CHAR |
| 7. | BLK | 1998 BLOCK AND SUFFIX | 6 | 56 - 61 | CHAR |
| 8. | URBNZ | URBANIZATION | 30 | 62 - 91 | CHAR |
| 9. | HSNUM | HOUSE NUMBER (LJ/BF) | 10 | 92 - 101 | CHAR |
| 10. | SNAME | STREET NAME (LJ/BF) | 35 | 102 - 136 | CHAR |
| 11. | UNIT | UNIT DESIGNATION (LJ/BF) | 15 | 137 - 151 | CHAR |
| 12. | RR | RURAL ROUTE/BOX # (LJ/BF) | 25 | 152 - 176 | CHAR |
| 13. | POBX | PO BOX NUMBER (LJ/BF) | 10 | 177 - 186 | CHAR |
| 14. | CITY | CITY/TOWN NAME | 20 | 187 - 206 | CHAR |
| 15. | ZIP | ZIP CODE | 5 | 207 - 211 | CHAR |
| 16. | ZIP4 | ZIP + 4 | 4 | 212 - 215 | CHAR |
| 17. | STATE | FIPS STATE ABBREVIATION | 2 | 216 - 217 | CHAR |
| 18. | FIPSCNTY | FIPS COUNTY CODE | 3 | 218 - 220 | CHAR |
| 19. | FIPST | FIPS STATE CODE | 2 | 221 - 222 | CHAR |
| 20. | PL | PHYSICAL LOCATION DESCRIPTION | 50 | 223 - 272 | CHAR |
| 21. | PRKNM | TRAILER PARK NAME | 30 | 273 - 302 | CHAR |
| 22. | HUFIN | MATCH CODE FROM HU MATCHING | 2 | 303 - 304 | CHAR |
| 23. | HUFINID | ID FROM HOUSING UNIT MATCHING | 12 | 305 - 316 | CHAR |
| 24. | TOA | TYPE OF BASIC ADDRESS<br>1 = ONE FAMILY HOUSE<br>2 = BSA WITH 2 OR MORE HUS<br>3 = MOBILE HOME NOT IN PARK<br>4 = MOBILE HOME IN PARK<br>5 = ONE FAMILY HOME IN SPECIAL PLACE<br>6 = BSA WITH 2 OR MORE HUS IN A SPECIAL PLACE<br>7 = OTHER | 1 | 317 - 317 | CHAR |
| 25. | USTAT | UNIT STATUS<br>1 = OCCUPIED OR VACANT AND INTENDED FOR OCCUPANCY<br>2 = UNDER CONSTRUCTION<br>3 = FUTURE CONSTRUCTION<br>4 = UNFIT FOR HABITATION<br>5 = BOARDED UP<br>6 = STORAGE OF HOUSEHOLD | 1 | 318 - 318 | CHAR |

Layout Name : ENHANCED00.LAY Page : 2
Description : 2000 ENHANCED LIST LAYOUT
Total Length : 360
Date Created : 11-22-1999

| # | Field | Field description | length | Positions Beg | - | End | |
|---|---|---|---|---|---|---|---|
| | | GOODS<br>7 = VACANT MOBILE HOME SITE<br>8 = OTHER | | | | | |
| 26. | UR | URBAN/RURAL<br>1 = URBAN<br>2 = RURAL | 1 | 319 | - | 319 | CHAR |
| 27. | QAFLG | QA SAMPLE FLAG<br>0 = NOT IN QA SAMPLE<br>1 = IN QA SAMPLE | 1 | 320 | - | 320 | CHAR |
| 28. | ESAMPFLG | E-SAMPLE ELIGIBILITY FLAG | 1 | 321 | - | 321 | CHAR |
| 29. | URFLAG | FLAG INDICATING THAT ADDRESS IS CONSIDERED TO BE URBAN OR RURAL<br>0 = RURAL<br>1 = URBAN | 1 | 322 | - | 322 | CHAR |
| 30. | MULTIFLAG | FLAG INDICATING THAT UNIT IS IN A MULTIUNIT OF LESS THAN 20 UNITS<br>0 = MULTI <20 UNITS<br>1 = NONMULTI, OR MULTI >= 20 | 1 | 323 | - | 323 | CHAR |
| 31. | DSSDSEG | SEGMENT FOR LARGE BLOCK SUBSAM | 2 | 324 | - | 325 | CHAR |
| 32. | FLDSEG | SEGMENT FOR ASSIGNING WORK IN | 2 | 326 | - | 327 | CHAR |
| 33. | INTERVW | AFTER LARGE BLOCK SUBSAMP<br>0 = OUT OF SAMPLE<br>1 = IN SAMPLE | 1 | 328 | - | 328 | CHAR |
| 34. | JIC | JUST IN CASE SPACE<br>**********<br>THESE FIELDS ARE USED FOR LARG BLOCK SUBSAMPLING.<br>********** | 10 | 329 | - | 338 | CHAR |
| 35. | UNITCNT | NUMBER OF UNITS IN STRUCTURE | 4 | 339 | - | 342 | CHAR |
| 36. | TOTCASES | NUMBER OF CASES IN CLUSTER | 6 | 343 | - | 348 | CHAR |
| 37. | ICMCASES | NUMBER OF ICM CASES IN CLUSTER | 6 | 349 | - | 354 | CHAR |
| 38. | CENCASES | NUMBER OF CEN CASES IN CLUSTER | 6 | 355 | - | 360 | CHAR |

# Sample Design File

The Sample Design File contains one record per block cluster selected during the listing sample selection. If the block cluster falls out of sample during the second step of the listing sample, the A.C.E. reduction, small block cluster subsampling, or the A.C.E. reduction, the remaining variables will be left blank. The initial version of the file, which will be created following the initial block cluster selection, is called SDF.US1. For each subsequent update to the file, the version number will increase by one (i.e. SDF.US2, SDF.US3). The layout for the Sample Design File is as follows:

| Variable Description | Name | Places | Source |
|---|---|---|---|
| Census Region | REGION | 1 | UN |
| Census Division | DIV | 2 | UN |
| State code | STATE | 3-4 | UN |
| County code | COUNTY | 5-7 | UN |
| Local census office | LCO | 8-11 | CS |
| Interim Tract (Pseudo Tract) | ITRACT | 12-17 | BC |
| Current Sample Indicator | CSI | 19 | UO |
| A.C.E. block cluster number | CLUST | 21-25 | CS |
| Check Digit | DIGIT | 26 | CS |
| Geography block cluster number | GCLUST | 28-32 | BC |
| List/Enumerate Indicator | LEIND | 33 | BC |
| Type of Enumeration Area Recode | TEACR | 34 | CS |
| Type of Enumeration Area group | TEAG | 36 | BC |
| Number of HUs used for sample design | NHU | 37-41 | BC |
| Number of MAF HUs | NHUM | 43-47 | BC |
| Number of 1990 HUs | NHU90 | 49-53 | BC |
| Sampling Stratum<br>1 = Small<br>2 = Medium<br>3 = Large<br>4 = American Indian Reservation | SS | 55 | UN |
| American Indian Country Indicator<br>0 = No American Indian Country<br>1 = American Indian Reservation/trust land<br>2 = Tribal Jurisdiction Area/<br>Alaska Native Village Statistical Area/<br>Tribal Designated Statistical Area | AICIND | 56 | BC |
| Demographic/Tenure Group code | DTCODE | 57-58 | UN |
| Demographic/Tenure Group label | DTLABEL | 59-60 | UN |
| Estimated Urbanicity of block cluster<br>1 = Urban Area with population ≥250,000<br>2 = Other Urban Area<br>3 = Non-Urban Area | ECLUSURB | 62 | UN |
| Size Category<br>1=Small (0-2 hus)<br>2=Medium (3-79 hus)<br>3=Large (80+ hus) | SIZCAT | 63 | UN |
| Additional space | | 64-91 | |

| Variable Description | Name | Places | Source |
|---|---|---|---|
| First step index number | INDEX1 | 92-99 | CS |
| Listing sample selection indicator<br>1 = Selected | BC1 | 101 | CS |
| Random Start for listing sample selection | RS1 | 103-113 | UN |
| Take-every for listing sample selection | TE1 | 115-125 | UN |
| Second step listing sample selection indicator<br>0 = Not Selected<br>1 = Selected | BC2 | 127 | CS |
| Random Start for the second step of the listing sampling | RS2 | 129-139 | CS |
| Take-every for the second step of the listing sampling | TE2 | 141-151 | CS |
| Unbiased weight after block cluster sampling | WEIGHTBC | 153-164 | CS |
| Additional space | | 165-175 | |
| Preliminary Number of HUs on the Independent List | NHUILP | 176-180 | AR |
| Number of Housing Units On the January 2000 DMAF | NHUDMAF | 182-186 | AR |
| Demographic Code<br>1 = Minority<br>2 = Non-Minority<br>3 = Puerto-Rico | DEMCODE | 188 | AR |
| Consistency Code<br>1 = Low Inconsistent (IL significantly smaller than DMAF)<br>2 = Consistent<br>3 = High Inconsistent ((IL significantly larger than DMAF) | CONCODE | 189 | AR |
| A.C.E. Reduction Stratum | ARST | 190-191 | AR |
| A.C.E. Reduction Indicator<br>0 = Not Selected<br>1 = Selected | ACERED | 193 | AR |
| Random Start for A.C.E. Reduction | RSAR | 195-205 | AR |
| Take-every for A.C.E. Reduction | TEAR | 207-217 | AR |
| Unbiased weight after A.C.E. reduction | WEIGHTAR | 219-230 | AR |
| Collapsing Flag | COLFLAG | 232 | AR |
| A.C.E. Reduction Index Number | INDEXR | 234-241 | AR |
| Number of Housing Units On the December 1999 DMAF (Initial) | NHUDMAFI | 243-247 | AR |
| Additional space | | 248-300 | |
| Number of HUs on the Independent List | NHUIL | 301-305 | SB |
| Small Block Cluster Subsampling Stratum | SBCSS | 306-307 | SB |
| Small Block Subsampling Indicator<br>0 = Not Selected<br>1 = Selected | SB | 308 | SB |
| Random Start for Small Block subsampling | RSSB | 310-320 | SB |
| Initial take-every for Small Block subsampling | ITESB | 322-332 | SB |
| Unbiased weight for A.C.E. cluster | WEIGHTC | 334-345 | SB |
| Larger of the DMAF and IL HU count | LARGERHU | 347-351 | SB |
| Final take-every for Small Block subsampling | FTESB | 352-362 | SB |
| Additional space | | 363-370 | |

| Variable Description | Name | Places | Source |
|---|---|---|---|
| Relisted Block Cluster Flag<br>0 = Not Relisted, 1 = Relisted | RELIST | 371 | LB |
| Number of total hus in block cluster | NHUEL | 373-377 | LB |
| Number of A.C.E. hus in cluster | NHUELA | 379-383 | LB |
| Number of supplemental hus in cluster | NHUELN | 385-389 | LB |
| Large Block Cluster EL subsampling code<br>1 = NHUELI< 80 hus, 2 = NHUELI ≥ 80 hus | ELLBSUB | 391 | LB |
| Random Start for Large Block subsampling | RSLB | 393-403 | LB |
| Take-every for Large Block subsampling | TELB | 405-415 | LB |
| Number of segments in block cluster | NSEG | 417-418 | LB |
| Number of segments selected in block cluster | NSEGSAM | 420-421 | LB |
| Day of Arrival | DAY | 423-424 | LB |
| Final Cluster Order Number | CON | 431-434 | LB |
| Number of total hus for interview in block cluster | NINT | 436-440 | LB |
| Unbiased weight for P-sample HUs | WEIGHTP | 442-453 | LB |
| Number of Assignments in block cluster | NA | 455-456 | LB |
| Final Sampling Strata | FSS | 458-464 | LB |
| Additional space | | 465-490 | |

| Variable Description | Name | Places | Source |
|---|---|---|---|
| Number of confirmed A.C.E. housing units not found in the census | CURCI | 676-680 | TES |
| Number of unconfirmed A.C.E. housing units not found in the census | CURUI | 682-686 | TES |
| Number of census housing units geocoded to the wrong census block | CURGE | 688-692 | TES |
| Targeted extended search selection type | TESSELECT | 694 | TES |
| Targeted extended search selection flag | TESFLAG | 696 | TES |
| Random Start for the targeted extended search | RSTES | 698–709 | TES |
| Take-every for the targeted extended search | TETES | 710–721 | TES |
| Targeted Extended Search Index Number | TESN | 722-727 | TES |
| Additional Space | | 728-750 | |

Source Codes

AR: A.C.E. Reduction
BC: Block Clustering
CS: Block Cluster Sampling
LB: Large Block Subsampling
SB: Small Block Subsampling
TES: Targeted Extended Search
UN: Universe File Creation
UO: Updated for each operation

Large Block Cluster Subsampling Example

This hypothetical example demonstrates the phases that a block cluster goes through during the large block cluster subsampling process.

1. Calculate the Sampling Parameters for Each A.C.E. Reduction Stratum and State (see reference 3)

   Sampling parameter calculations occur prior to the arrival of block clusters. This information is based on results from the IL. Let's say for a particular state and A.C.E. reduction stratum, the target number of HUs, T, the number of HUs in block clusters with more than 80 A.C.E. HUs, NILHUL, and the number of HUs in block clusters with 0-79 A.C.E. HUs (except small block clusters with less than 10 IL HUs), NILHUM, are:

   T = 2050
   NILHUL = 1295
   NILHUM = 1173

   The sampling parameters calculated from this information are the take-every, TELB, the number of segments per block cluster, NSEG, and the random start, RS.

$$TELB = \frac{NILHUL}{T - NILHUM} = \frac{1295}{2050 - 1173} = 1.477000$$

$$NSEG = \frac{1}{1 - \frac{1}{TELB}} = \frac{1}{1 - \frac{1}{1.477000}} = 3.096000 \text{ (Rounded up to the next integer)} = 4$$

   FORMULA = 2, since formula 2 was used.

A random number is selected, RN = 0.179317, and the random start is calculated.

$$RS = RN \times 1.477 = 0.179317 \times 1.477000 = 0.265000$$

The remaining phases in large block cluster subsampling rely on information obtained from the HUs on the preliminary EL. Suppose two block clusters from the same A.C.E. reduction stratum within a state arrive on day one. These block clusters go through large block cluster processing as follows:

2. Create Block Cluster Housing Unit Counts and Subsampling Status Codes

To determine whether each block cluster needs to be segmented and subsampled or the entire cluster remains in sample, HU counts are calculated. The two types of HUs are A.C.E. HUs, NHUELA, and supplemental HUs, NHUELN. The combination of these two types is the total HUs on the preliminary EL, NHUEL. The subsampling codes, ELLBSUB, SEGSUB, SEGSAM, and SEGID are assigned to the block clusters that do not need to be segmented and subsampled.

| Block Cluster | NHUEL | NHUELA | NHUELN | ELLBSUB | SEGSUB | SEGSAM | SEGID | In AIR? | Next Phase |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 247 | 197 | 50 | 1 | 1 | 1 | AA | yes | Send to 6 |
| 2 | 83 | 82 | 1 | 2 | TBD | TBD | TBD | no | Send to 3 |

3. Create Segments

For those non-AIR block clusters with 80 or more A.C.E. HUs (block cluster 2), segments need to be formed. The most that the segments within a block cluster will differ is one A.C.E. HU. A segment identifier is also assigned to distinguish between segments in each block cluster.

$$AVGSEG = \text{trunc}\left(\frac{NHUELA}{NSEG}\right)$$

$$R = NHUELA - NSEG \times AVGSEG$$

The supplemental HUs are assigned to the same segment as the preceding A.C.E. HU based on map spot number. For this example, suppose the one supplemental HU is assigned to the segment CA.

| Block Cluster | Avg. | AVGSEG | R | Segment AA A.C.E. HUs (Supplemental HUs) | Segment BA A.C.E. HUs (Supplemental HUs) | Segment CA A.C.E. HUs (Supplemental HUs) | Segment DA A.C.E. HUs (Supplemental HUs) |
|---|---|---|---|---|---|---|---|
| 2 | 20.5 | 20 | 2 | 21 (0) | 21 (0) | 20 (1) | 20 (0) |

4. Create Segment Level Variables and Codes

HU counts were calculated previously for block clusters. In this section, HUs are counted similarly for each segment. A.C.E. HUs, supplemental HUs, and total HUs are counted.

| Block Cluster | SEGID | NHUELS | NHUELAS | NHUELNS |
|---|---|---|---|---|
| 2 | AA | 21 | 21 | 0 |
| | BA | 21 | 21 | 0 |
| | CA | 21 | 20 | 1 |
| | DA | 20 | 20 | 0 |

5. Select a Systematic Sample of Segments for Each A.C.E. Reduction Stratum and State

After the non-AIR block clusters with 80 or more HUs have been segmented, a systematic sample of segments is selected using the parameters calculated in step 1. The sampling is across all segments in all block clusters within the same A.C.E. reduction stratum, state, and day. Since the selection is done over several days, the systematic sample needs to be carried over from day to day. A daily segment order number (DSON) is assigned to implement the sampling on a daily basis. The cluster order number (CON) over all days is also assigned. Information to be carried over are the daily end value (DE), which is the following day's start value (DS), and the cumulative cluster count (CCC), which is a cumulative count of block clusters processed.

So for day 1:

TE = 1.477000 and DS = 0.265000

L1 = DS = 0.26500 → take segment 1
L2 = 1.742000 → take segment 2
L3 = 3.219000 → take segment 4

| Day | CON | SEGID | DSON | SEGSUB |
|---|---|---|---|---|
| 1 | 1 | AA | 1 | 1 |
| | | BA | 2 | 1 |
| | | CA | 3 | 0 |
| | | DA | 4 | 1 |

N = 4 and n = 3
DE = 0.265000 + 1.477000×3 - 4 = 0.696000 →start day 2 with a DS = 0.696000
CCC = 1 →start day 2 with a CON = CCC + 1 = 2

6. Identify Housing Units for A.C.E. Interview

If an A.C.E. HU has an INTERVW of 1, then that HU is sent for interview. If an A.C.E. HU has an INTERVW of 0, then that HU is not sent for interview. Supplemental HUs have an INTERVW of 8 or 9 and none of them will be sent for interview.

| Block Cluster | Segment | INTERVW | Result |
|---|---|---|---|
| 1 | AA | 1 for A.C.E. HUs<br>9 for supp. HUs | Interview 197 A.C.E. HUs |
| 2 | AA | 1 for A.C.E. HUs | Interview 21 A.C.E. HUs |
| | BA | 1 for A.C.E. HUs | Interview 21 A.C.E. HUs |
| | CA | 0 for A.C.E. HUs<br>8 for supp. HU | Interview 0 A.C.E. HUs |
| | DA | 1 for A.C.E. HUs | Interview 20 A.C.E. HUs |

7. Create Interviewer Workload Assignments

After determining the number of HUs to be sent for interview in a block cluster, the next step is to determine the number of assignments. A block cluster with 80 or more HUs to be sent for interview is divided into two or more assignments of about 40 to 50 HUs per assignment. Block clusters with 0-79 HUs to be sent for interview are left as one assignment. The number of HUs for interview is 197 in block cluster 1 and 62 in block cluster 2, thus only block cluster 1 needs to be split into assignments.

$$NA = \frac{NINT}{40} = \frac{197}{40} = 4.9 \text{ (Rounded down)} = 4 \text{ Assignments}$$

$$AVGHUA = \text{trunc}(\frac{NINT}{NA}) = \text{trunc}(\frac{197}{4}) = 49$$

$$RINT = NINT - NA \times AVGHUA = 197 - 4 \times 49 = 1$$

Give the first assignment one additional HU than the last three assignments.

For the first assignment:
ASIZE = 49 + 1 = 50 interviews
For the next three assignments:
ASIZE = 49 interviews

To distinguish between assignments within a block cluster ASSIGNID is assigned. For block cluster 2, ASSIGNID is set to AA since it did not require multiple workload assignments.

| Block Cluster | Assignment | ASSIGNID | Interviews |
|---|---|---|---|
| 1 | 1 | AA | 50 |
| | 2 | AB | 49 |
| | 3 | AC | 49 |
| | 4 | AD | 49 |
| 2 | 1 | AA | 62 |

# Overview of the Large Block Cluster Subsampling Process